Genomics 92 (2008) 478–487

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm

I.Y. Abdurakhmonov [a,*], R.J. Kohel [b,1], J.Z. Yu [b,1], A.E. Pepper [c,2], A.A. Abdullaev [a], F.N. Kushanov [a], I.B. Salakhutdinov [a], Z.T. Buriev [a], S. Saha [d,3], B.E. Scheffler [e], J.N. Jenkins [d], A. Abdukarimov [a]

[a] *Center of Genomic Technologies, Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan. Yuqori Yuz, Qibray region Tashkent district, 702151, Uzbekistan*
[b] *USDA-ARS, Crop Germplasm Research Unit, 2881 F&B Road College Station, TX 77845, USA*
[c] *Department of Biology, Texas A&M University, College Station, Texas 77843, USA*
[d] *USDA-ARS, Crop Science Research Laboratory, Genetics and Precision Agriculture, P. O. Box 5367, 8120 Highway 12E, Mississippi State, MS 39762, USA*
[e] *USDA-ARS, 141 Experiment Station Road, Stoneville, MS 38776, USA*

ARTICLE INFO

Article history:
Received 29 January 2008
Accepted 29 July 2008
Available online 5 October 2008

Keywords:
Cotton germplasm
Genetic diversity
Fiber quality
Linkage disequilibrium (LD)
Simple sequence repeats (SSR) markers
LD-based association mapping

ABSTRACT

The narrow genetic base of cultivated cotton germplasm is hindering the cotton productivity worldwide. Although potential genetic diversity exists in *Gossypium* genus, it is largely 'underutilized' due to photoperiodism and the lack of innovative tools to overcome such challenges. The application of linkage disequilibrium (LD)-based association mapping is an alternative powerful molecular tool to dissect and exploit the natural genetic diversity conserved within cotton germplasm collections, greatly accelerating still 'lagging' cotton marker-assisted selection (MAS) programs. However, the extent of genome-wide linkage disequilibrium (LD) has not been determined in cotton. We report the extent of genome-wide LD and association mapping of fiber quality traits by using a 95 core set of microsatellite markers in a total of 285 exotic *Gossypium hirsutum* accessions, comprising of 208 landrace stocks and 77 photoperiodic variety accessions. We demonstrated the existence of useful genetic diversity within exotic cotton germplasm. In this germplasm set, 11–12% of SSR loci pairs revealed a significant LD. At the significance threshold ($r^2 \geq 0.1$), a genome-wide average of LD declines within the genetic distance at <10 cM in the landrace stocks germplasm and >30 cM in variety germplasm. Genome wide LD at $r^2 \geq 0.2$ was reduced on average to ~1–2 cM in the landrace stock germplasm and 6–8 cM in variety germplasm, providing evidence of the potential for association mapping of agronomically important traits in cotton. We observed significant population structure and relatedness in assayed germplasm. Consequently, the application of the mixed liner model (MLM), considering both kinship (K) and population structure (Q) detected between 6% and 13% of SSR markers associated with the main fiber quality traits in cotton. Our results highlight for the first time the feasibility and potential of association mapping, with consideration of the population structure and stratification existing in cotton germplasm resources. The number of SSR markers associated with fiber quality traits in diverse cotton germplasm, which broadly covered many historical meiotic events, should be useful to effectively exploit potentially new genetic variation by using MAS programs.

Introduction

The genus *Gossypium* includes approximately 45 diploid and 5 allotetraploid species distributed mostly in tropical and subtropical regions of the world [1,2]. It provides most of the world's natural textile fiber, sources of oil, and cottonseed meal as feed products [3,4]. Diploid cottons are classified into eight (A–G to K) cytogenetically defined genome groups [5]. Hybridization between A-genome and D-genome diploids and subsequent polyploidization about 1.5 million years ago created the five AD allotetraploid lineages that are indigenous to the Americas and Hawaii [6–8]. These New World tetraploid cottons include the commercially important species, *Gossypium hirsutum* and *G. barbadense*, where *G. hirsutum* is the most widely cultivated (90%) industrial cotton among all *Gossypium* species [9,10]. The origin of *G. hirsutum* is Guatemala, but its large indigenous range encompassed most of Mesoamerica and Caribbean. According to archaeobotanical findings, *G hirsutum* probably was domesticated originally within the Southern end of Mesoamerican gene pool [11,12]. Mexico–Guatemala is considered the site of original domestication and primary center of diversity. *G. hirsutum* exhibits a diversity of types including a single wild race, 'yucatanense', six primitive domesticated forms, 'richmondi', 'punctatum', 'latifolium', 'palmeri',

* Corresponding author. Fax: +7 998 712 642 390.
E-mail address: genomics@uzsci.net (I.Y. Abdurakhmonov).
[1] Phone: +1 979 260 9311; fax: +1 979 260 9333.
[2] Phone: +1 979 845 2518; fax: +1 979 862 4790.
[3] Phone: +1 662 320 7386; fax: +1 662 320 7528.

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.07.013

Table 1
Summary of fiber quality traits from Mexican environment

Traits	Number	Mean	Minimum	1st Q	median	3rd Q	Maximum
Micronaire (MIC)	203	4.73	3	4.25	4.7	5.15	6.6
Fiber length (UHM)	203	1.09	0.84	1.04	1.1	1.15	1.28
Uniformity (UI)	203	83.84	79	82.45	83.8	85.35	88
Strength (STR)	203	27.54	21.3	25.55	27.4	29.3	35.5
Elongation (ELO)	203	6.52	3.8	5.8	6.5	7.2	10.1
Reflectance (RD)	203	72.11	0	74.65	76.2	77.9	81.5

Note: 1st Q–25%-ile; 3rd Q–75%-ile.

'morilli', 'marie-galante' stages [13,14], and a number of domesticated variety accessions, 'euhirsutum' [15], grown in 80 different countries worldwide [3]. The accessions of wild and primitive races of *G. hirsutum* are the 'golden' reservoirs of genetic diversity, but they are largely 'underutilized' in breeding programs due to their short-day photoperiodic flowering [16,17]. The 'euhirsutum' accessions are the main germplasm resources of worldwide cotton breeding programs to improve cultivars, and world cotton production greatly relies on these variety germplasm. Consequently, these variety germplasm of *G. hirsutum* make the major portion (at least 50–80%) of cotton germplasm collections worldwide [10]. Currently, cotton researchers and producers are concerned with the narrow genetic base of cultivated cottons that has experienced recent cotton yield and quality declines [18,19]. The yield and quality stagnation could be directly attributed to many factors, perhaps, including rapid genetic erosion of genetically uniform elite cultivars and paucity of information on the complex nature of the cotton genome especially in relation to economically important traits. These indicate the critical need to 1) explore novel germplasm resources for potential natural genetic diversity and 2) develop innovative genomics tools to efficiently mobilize these useful genetic variations to breeding germplasm. This should help to overcome existing and potential problems of worldwide cotton production associated with narrow genetic base of the cultivar germplasm.

During past years, the international cotton research community has developed extensive genomic resources [20,21]. A large collection of robust, portable, and PCR-based Simple Sequence Repeat (SSR) marker resources were developed in multiple laboratories and made available to the cotton research community through the cotton marker database (CMD) [22]. Molecular marker technology was successfully used to assess genetic diversity of germplasm resources, create genetic linkage maps, and map important agronomic QTLs in bi-parental mapping populations [20,21], which are imperative for acceleration of marker-assisted-selection (MAS) programs in cotton. Alternatively, turning the gene-tagging efforts from bi-parental crosses to germplasm collections, and from traditional linkage mapping to linkage disequilibrium (LD)-based association study promises the most effective utilization of *ex situ* conserved natural genetic diversity of worldwide cotton germplasm resources. LD refers to a historically reduced (non-equilibrium) level of the recombination of specific alleles at different loci controlling particular genetic variation in a population. This LD can be measured statistically, and has been widely applied to map and eventually clone a number of genes underlying complex genetic traits in humans [23,24].

The advantages of population-based association study over conventional QTL mapping in bi-parental crosses primarily are due to 1) availability of broader genetic variations with wider background for marker-trait correlations (i.e. multiple alleles evaluated simultaneously), 2) likelihood of a higher resolution mapping because of the utilization of majority recombination events in the germplasm's developmental history, 3) possibility of exploiting historically measured trait data for associations, and 4) no need for the development of expensive and tedious bi-parental populations that makes the approach timesaving and cost-effective [25–27] . Conversely, traditional QTL mapping is 1) very costly [27], 2) has poor resolution with the evaluation of only a few alleles [28], and 3) it requires a longer research time period. Additionally, association mapping in plants, compared to human populations, has more power with the opportunity to create mapping populations with required amount of LD and diversity [29].

The measurement of the LD patterns for genomic regions and specificity of LD extent among different populations or groups of the 'target' organisms is the important starting point to design and conduct association mapping [30,31]. LD has been quantified in several plant species [32,33] including the model organism *Arabidopsis* and now extended to crops such as maize, barley, durum wheat, spring wheat, sorghum, soybean, sugarcane, sugar beet and grapevine, as well as in trees such as European aspen, and loblolly pine [28,32,33]. These studies revealed that the genome-wide extent of LD varied across genomes and between species with the examples of longer stretches of LD in some local populations. Moreover, LD-based association mapping was successfully used in plant germplasm resources that highlighted serious influence of population structure and relatedness of individuals in conducting association mapping [32,33].

The application of LD-based association mapping for cotton will not only accelerate still lagging MAS programs in cotton, but also adds to our knowledge and understanding of the complex cotton genome and its evolution. However, the extent of genome-wide LD has not been determined in cotton. In this study, we utilized largely 'unexploited' cotton germplasm resource from the Republic of Uzbekistan to search for new genetic variation in fiber quality traits in cotton and mapping the main fiber quality traits using LD-based association mapping strategy. Here we report the extent and distribution of molecular diversity, population structure, kinship, and an average extent of genome-wide LD for exotic germplasm measured using SSR markers. LD-based association mapping for fiber quality traits was applied using mixed linear model approach that found several SSR markers associated with the main fiber quality traits of exotic cotton germplasm. The results provide preliminary insight into genome-wide averages for extent of LD in cotton and are very useful as a framework for future 'association studies' in cotton that will accelerate development of superior cotton cultivars through MAS programs.

Results

Fiber quality properties of selected accessions in Mexican environment

The cotton accessions including 208 landrace stocks and 77 photoperiodic variety accessions from Mexico and Africa (Table S1) revealed a wide range of phenotypic variation among fiber quality traits including fiber strength (STR), fiber length (upper half mean–UHM), uniformity (UI), elongation (ELO), micronaire (MIC), and reflectance (Rd) (Table 1). We observed significant trait correlations among different fiber traits in the Mexican environment (Tables 2). The negative correlations were observed between MIC and UHM, MIC and STR, and STR and ELO. Positive correlations were observed

Table 2
Correlation of fiber quality traits from Mexican environment

TRAITS	MIC	UHM	UI	STR	ELO	RD
MIC	1					
UHM	−0.49****	1				
UI	−0.09	0.44****	1			
STR	−0.32****	0.69****	0.29****	1		
ELO	−0.02	−0.11	0.09	−0.29****	1	
RD	0.09	0.28****	0.22**	0.32****	−0.03	1

MIC—Micronaire; UHM—fiber length; UI—uniformity; STR—fiber strength; ELO—elongation; RD—reflectance. **, ****, $p \leq 0.01$, 0.001, respectively.

Table 3
Summary of SSR polymorphisms

Accession panels	No. of taxa	No. of polymorphic SSRs				Polymorphic Information Content (PIC)	
		Overall	Unique (%)	Rare (%)	Average allele/locus	Range	Average
Exotic panel	287*	373	3	49	4	0.007–0.38	0.122
Exotic landraces only	208**	370	3	43	4	0.01–0.38	0.134
Mexican and African variety group	78**	161	0.6	36	2	0.02–0.37	0.160

* panel included *G. hirsutum* (TM-1) and *G. barbadense* (3–79) controls.
** panel included only *G. hirsutum* (TM-1) control.

between UHM and UI, UHM and STR, UHM and Rd, UI and Rd, and STR and Rd (Table 2).

SSR polymorphism

In 287 exotic cotton accessions, including 208 landrace stocks, 77 photoperiodic variety accessions from Mexico and Africa, and control genotypes of Texas Marker-1 and Pima 3-79 (Table S1), 95 primer pairs (an average of 4 SSR primer pairs per each chromosome, covering ~18% of the cotton genome; Table S2) detected 373 polymorphic SSR alleles or amplicons with an average of 4 SSR (a range of 2–8 alleles) alleles per primer pair. We found 10 (~3%) unique SSR alleles that were specific to 0.5% of the accessions. The 182 SSR (~49%) alleles were rare and present in only 5% of the cotton accessions. The remaining 181 (48%) SSRs were highly polymorphic (Table 3). The overall polymorphic information content (PIC) for SSR markers was in a range of 0.007–0.380 with an average of 0.122, but we observed an average increase in PIC values in separate analyses of landrace stocks (0.134) and variety germplasm group of Mexican and African origin (0.160). The average similar frequency of alternative alleles of SSR data was 0.42 for dominant versus 0.55 for recessive alleles, after removing 5% minor frequency alleles. Filtering of original SSR data set for a 5% minor allele frequency generated SSR data set from 70 SSR primer pairs, which covered ~12% of the cotton genome with an average of ~3 SSR primer pairs per each cotton chromosome spanning at an average of ~46 cM distance/chromosome (Table S2).

Genetic distance estimates

The overall NJ-analysis (Fig. 1) revealed genetic distance (GD) among all *G. hirsutum* accessions ranged from 0.01–0.50 with average of 0.13 demonstrating significant genetic diversity ranges. The overall GD range among only cotton landrace stocks was equal to 0.02–0.50 with the average distance of 0.26. The lower overall genetic distances were observed within a group of Mexican and African variety cotton accessions (average of 0.07–0.08), revealing the narrow genetic base of cultivated *G. hirsutum* cottons of these two diverse ecotypes. The average GD between TM-1 and all cotton accessions was 0.08–0.12. However, we observed several landrace stocks (belonging to ssp.

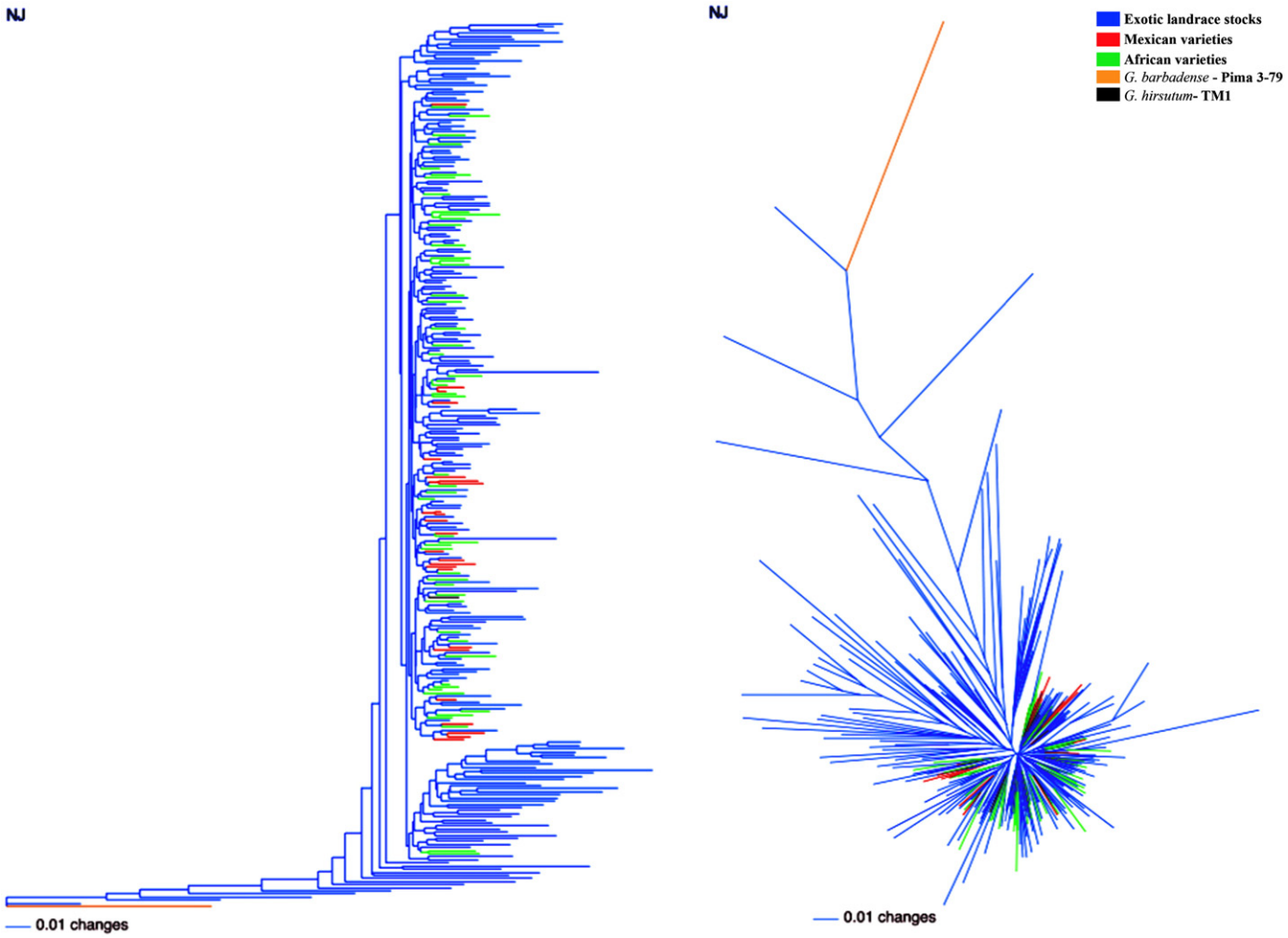

Fig. 1. Rooted and Unrooted Neighbor-joining (NJ) trees; the control lines, ecotypes, and landrace stocks are color coded for simplicity. Branch length is shown.

Table 4
Summary of genetic distances (GD) from PAUP analysis

Accession panels	Overall GD		GD with *G. hirsutum* (TM-1)		GD with *G. barbadense* (3–79)		No. of accessions
	Range	Average	Range	Average	Range	Average	
Exotic panel (All)	0.01–0.50	0.13	0.03–0.43	0.10	0.22–0.62	0.54	285
Exotic landraces	0.02–0.50	0.26	0.05–0.43	0.12	0.22–0.62	0.54	208
Mexican varieties	0.01–0.12	0.07	0.06–0.11	0.08	0.53–0.56	0.55	24
African varieties	0.03–0.15	0.08	0.04–0.12	0.08	0.51–0.57	0.54	53
TM-1	–	–	–	–	0.54	–	

mexicanum and punctatum) with the highest GD of 0.27–0.43 from TM-1, demonstrating existence of wider genetic diversity in *G. hirsutum* cotton landrace stocks. An average GD between cotton accessions and *G. barbadense* standard 3–79 was 0.54 with the highest distance (GD=0.62) from *G. hirsutum* landrace stocks. The GD between TM-1 and 3–79 was equal to 0.54 when genotyped with the 95 core set SSR markers (Table 4). The genetic base of variety cotton accessions seems narrower than landrace stock germplasm, demonstrating the potential for landrace stock accessions to contain wider genetic diversity.

Analysis of molecular variance (AMOVA) in cotton accession groups

To estimate genetic diversity within and among groups, we analyzed Wright's F_{ST} index for predefined cotton accession groups using AMOVA test. The within group component of genetic variance prevailed and was attributed to 96.73% of the total variance. The 3.27% of the genetic variance was observed among groups with a significant overall F_{ST}=0.03274 (p<0.0001) (Table 5). Based on pairwise F_{ST} estimates between the predefined groups (Table 6), the African variety germplasm was closer to a group of landrace stocks as compared to the Mexican varieties, indicating possible extensive utilization of putative ancestors from landrace stocks in the development history of the African varieties. Population differentiation estimates (F_{ST}) between the predefined cotton germplasm groups in both panels were also analyzed using a Bayesian approach for dominant markers without prior knowledge of inbreeding coefficients. Several Bayesian models were run and the smallest deviance information criterion (DIC=2507.85) was observed with full model. The results of the $\theta^{(II)}$ (analogous to Weir and Cockerham's F_{ST}) test that corresponds to the amount of genetic differentiation among groups revealed the low level of genetic distance among the predefined cotton accession groups ($\theta^{(II)}$=0.06±0.006). The Bayesian analog of Nei's G_{ST} (G_{ST}-B) [34] also confirmed the low level of population differentiation (6%) among the predefined groups (data not shown). Although the greatest proportion of genetic variance of cotton germplasm groups was attributed to within population groups, those small variations observed among predefined groups were highly significant (p≤0.0001), suggesting the existence of population structure.

Assignments of cotton accessions into subgroups and kinship estimates

We further tested possible genetic clusters in these cotton germplasms using a model-based approach since information on the memberships of individuals to specific clusters (Q-matrix) and the relatedness of individuals (K-matrix) are crucial when conducting LD-based association mapping [35–37]. The model-based approach using prior population information (K=3) revealed shared ancestry information among groups (Fig. 2). All 208 landrace stock accessions were assigned to cluster 1 with a minimum probability of 0.57. According to posterior probability, 65 accessions in the cluster 1 shared recent ancestry with cluster 3 (African varieties) and five cotton accessions shared the ancestry from the Mexican cluster (cluster 2), whereas 22 landrace stock accessions shared recent ancestry with both the African and Mexican groups (Fig. 2). All 24 Mexican variety cotton accessions were assigned to cluster 2 with a minimum probability of 0.58. In this Mexican variety cotton cluster, three cotton accessions shared a substantial recent ancestry with the African cluster, and one cotton accession shared ancestry with both the landrace stock and African clusters. In the third African variety cotton cluster, all 53 African variety cotton accessions were assigned to cluster 3 with a minimum probability of 0.65, and only two cotton accessions from the Mexican cluster shared a significant posterior probability of being in this cluster. Results demonstrated that the African cluster accessions shared a greater genetic similarity to the exotic landrace stocks than do the accessions in the Mexican cluster.

Table 5
Analysis of molecular variance (AMOVA)

Source of variation	*df*	Sum of squares	Variance components	Percentage of variation	*p*-value
Among populations	2	124.450	0.685331	3.27	<0.0001
Within populations	283	5728.956	20.24366	96.73	<0.0001
Total	285	5853.406	20.928997		

Table 6
Pairwise and population specific F_{ST} estimated based on Weir and Cockerham approach

Groups	Landrace stocks	Mexican	African
Landrace stocks (208)	0.031****		
Mexican (24)	0.043****	**0.039****	
African (54)	0.026****	0.042****	**0.037****

Diagonal elements are population specific F_{ST}.
Below diagonal elements are pairwise F_{ST}.
**** Significant at p<0.0001.

The pairwise kinship values varied 0–1.0. We observed that the majority of the pairs of cotton accessions (55%) in both panels had zero estimated kinship values, while 22–23% of the pairs had a value of 0.05, and 20% of the pairs had a value 0.1–0.20. The remaining pairs of accessions (1–2%) had>0.25 kinship values, suggesting involvement of some common parental genotypes in these germplasm groups.

Pairwise linkage disequilibrium and LD decay

At significant threshold values (r^2≥0.05 and p≤0.005), 11–12% of the SSR marker pairs showed significant pairwise LD in a total 286 accessions of *G. hirsutum* (>15,753 pairwise comparisons). At the highly significant threshold of r^2≥0.1, only 4% of SSR marker pairs were remained in LD. In a separate analysis of 208 landrace stocks and 78 variety accessions of Mexican and African origin, we found a slightly different LD level between SSR markers. In the landrace stocks germplasm group 5% (r^2≥0.1) of SSR marker pairs (in 19, 701 pairwise comparisons) were in significant LD, whereas only 4% (r^2≥0.1) of the SSR marker pairs (in 5, 243 pairwise comparisons) showed significant LD in a group of accessions from Mexican and African ecotypes. This difference is larger (at p≤0.005, 3% in variety and 11% in landrace stocks) if LD is considered according to significant p-values (Table 7). It should be noted that the sample size of 78 for the variety germplasm of the Mexican and African group was small and results of pairwise LD might be biased. However, this difference could also be due to low polymorphism level of variety germplasm that was unable to efficiently detect LD in assayed regions.

We also tried to determine the structure of haplotypic LD in the genome since a strong block-like LD structure simplifies LD mapping of complex traits [38]. Triangle plots for pairwise LD between SSR markers demonstrated significant LD blocks in the genome-wide LD analysis (Fig. S1). These LD blocks are very useful in association mapping when sizes are calculated. Therefore, we tried to identify the

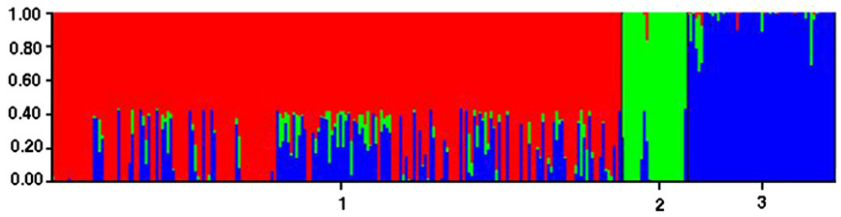

Fig. 2. The summary plot of Q-matrix estimates: cluster 1-landrace stock germplasm group (defined with the red color); cluster 2-Mexican varieties (defined with the green color); cluster 3-African varieties (defined with the blue color).

general sizes of these blocks in the cotton genome. This is called a genome-wide LD decay, in which r^2 LD values are plotted as a function of genetic distance in cM. We observed a significant ($r^2=0.1$) LD between pairs of SSR loci within 36–37 cM distance in the both landrace stock (Fig. 3A) and photoperiodic variety germplasm (Fig. 3B) groups, along with a few long stretched LD at 40–70 cM in landrace stock germplasm. However, LD clearly decays within the genetic distance of <10 cM in the landrace stocks germplasm with $r^2\geq0.1$, and >30 cM in Mexican and African variety germplasm group with $r^2\geq0.1$. This indicates that linkage is the main factor in conserving LD in cotton genome as is the case in many other organisms [27,33]. Genome wide LD at $r^2\geq0.2$ reduced to ~1–2 cM in the landrace stock germplasm group or 6–8 cM in variety germplasm group of Mexican and African ecotype. We observed significant LD between unlinked markers (Fig. 3), suggesting the existence of LD generating factors other than linkage in the cotton genome.

Association mapping of fiber quality traits

We performed association mapping of SSR loci with fiber quality (MIC, UHM, UI, STR, ELO, Rd) traits from the Cotton Winter Nursery-grown cotton accessions by using a Q+K mixed linear model (MLM) implemented in TASSEL. Out of 178 highly polymorphic SSR loci used for association mapping (i.e. a dataset with a 5% minor alleles filtered), 17 (~10%) SSRs were associated with MIC, 23 (13%) were associated with UHM, 18 (10%) were associated with UI, 19 (~11%) were associated with STR, 11 (6%) were associated with ELO, and 15 (8%) were associated with Rd traits (Fig. 4). These SSR markers, significantly associated with fiber quality traits in MLM, explained between 1% and 5% phenotypic variation, and MLM models in fiber trait associations explained between 3% and 53% phenotypic variations (data not shown) in all assayed fiber traits.

We compared fiber trait-associated SSR markers from our study with reported SSR markers from QTL-mapping analyses in various experimental populations [39–49]. Other published linkage mapping or association studies were not compared because of the use of a different marker system, or when compared, we did not mention them here due to not finding common SSR markers in our results. Thorough analysis of the literature data revealed that between 6% and 21% of the SSRs (or their alleles) showing association with the main fiber quality traits also were reported to be associated with fiber quality traits in other linkage-mapping studies in cotton (Fig. 4; Table S3). This supports the LD-based association mapping results [25] of our study that used diverse sets of cotton germplasm resources. The remaining SSRs are new markers revealing associations with fiber quality traits in diverse set of cotton germplasm. The details of association mapping results for each trait are listed in Table S3. We also checked whether SSR markers showing significant associations in diverse environments are in pairwise LD or not. The comparison of LD values (r^2) and association results from MLM ($p\leq0.05$) revealed that 85% of the SSRs significantly associated with fiber quality traits in MLM were in significant LD with other SSR loci at $r\geq0.1$ (Fig. S2).

Discussion

Our results demonstrated that the amplified SSR amplicons per primer pairs in assayed cotton accessions were lower (between 2 and 8 markers/primer pair) than that reported for landrace stocks [14,16,50]. The latter might be due to the use of the selected core set SSR markers in the genotyping. The genetic diversity studies of *Gossypium* species, inferred from isozymes, random amplified polymorphisms (RAPDs), restricted fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), and SSRs reported a low level of molecular diversity within *G. hirsutum* cotton germplasms (see review [50]). Our results obtained from genetic distance analysis confirmed the narrow genetic base of *G. hirsutum* cotton variety accessions. We found more genetic diversity among the exotic landrace stocks than breeding variety accessions of Mexico and Africa. Previous studies [14,16,50] also found wider genetic diversity in the landrace stocks of *G. hirsutum*, confirming the existence of sufficient genetic diversity in the exotic germplasm that is underutilized in current breeding programs.

A wide range of diversity in fiber quality traits of the exotic accessions also demonstrates the existence of potential genetic variation for primary fiber quality traits within *G. hirsutum* germplasm that is useful for future breeding programs. It is noteworthy to mention that an accurate quantitative trait analysis and tagging of functional QTLs require the phenotypic measurements in the multiple environments/locations. However, we were limited to use only one environment single replication trait data in our study primarily 1) due to the photoperiodic short-day flowering characteristics of assayed germplasms that are not productive in long day summer cultivations in both Uzbekistan and the U.S., and 2) high cost of growing these accessions in short-day conditions of the Cotton Winter Nursery at the Mexico for multiple years to get valuable trait estimates from breeding point of view. Furthermore, fiber quality traits analyzed in our study were reported to have high heritability values [e.g. [51]]. Although not perfectly ideal, this is the first time effort in phenotypic measurements of fiber quality traits of exotic accessions in such large samples, which allowed us to tag agronomically important QTLs of exotic cottons by using an association mapping strategy.

Table 7
The pairwise genome-wide linkage disequilibrium (LD) between pairs of SSR markers

Accessions	No. of accessions	No. of sites	No. of pairwise comparisons	LD, (%)				
				$p\leq0.05$	$p\leq0.01$	$p\leq0.005$	$r^2\geq0.05$	$r^2\geq0.1$
Exotics (all)	286	178	15753	25	17	12	11	4.0
Wild land races	208	199	19701	23	15	11	12	5.0
Varieties from Mexico and Africa	78	103	5243	7	4	3	11	4.0

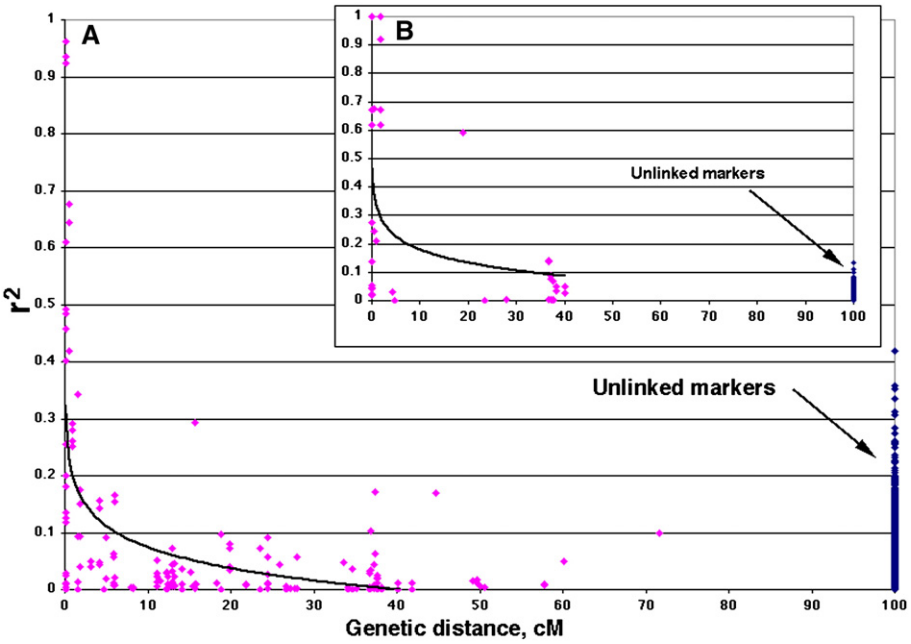

Fig. 3. LD decays within a distance: A landrace germplasm stocks and B variety germplasm group from Mexico and Africa. Inner fitted trend line is a non-linear logarithmic regression curve of r^2 on genetic distance. LD decay is considered below $r^2=0.1$ threshold based on trend line [86].

Insights to linkage disequilibrium in cotton genome

Majority of SSR primer pairs yielded multiple PCR-products in our cotton allotetraploid accessions. In diploid cotton, Ma et al. [52] reported that a large portion of expressed SSRs (eSSRs) yielded dominant amplification products in A-genome specific experimental mapping populations, where 77% of amplified eSSRs loci scored as dominant marker type. There is a great risk of false allele calling for multiple-band SSR markers when wide allotetraploid germplasm resources with unknown pedigree information are genotyped [33]. Misassignment of alleles is the concern in association analysis [53].

In spite of limited statistical power of dominant markers compared to co-dominant markers, they can be successfully applied to cluster populations using a Bayesian approach [35,54], estimate the kinship coefficients between individuals [55], quantify the genome-wide LD and conduct a LD-based association mapping in plants [25,26,33,56–58] provided a large sample size is used. Accordingly, the large sample size and the large number of heterozygous SSR loci genotyped in this study should give unbiased estimates of the genetic distance, genome-wide LD, population structure, and kinship with our dominantly scored SSR data.

Our results on allopolyploid cotton germplasm revealed the first insight into a level of pairwise LD for cotton measured with SSR markers. The percentage of SSR loci pairs in LD observed in cotton (11–12%) was comparable with reports in maize (10%) [59] and sorghum (8.7%) [60], yet was comparatively lower than that obtained in other studies. In different maize population groups, 49–56% of the SSR pairs were in significant LD [27,61]. Also a high percentage of SSR pairs in LD was reported for population groups of cultivated barley (45–100%) germplasm [25,62] as well as for the durum wheat (52–86%) elite germplasm [63]. A high recombination rate in allopolyploid cottons was reported [12] and it might be one of the factors explaining the observed low level of pairwise LD in cotton, along with mutation, selection, and genetic drift that occurred in the domestication of *G. hirsutum* germplasm.

Several reports in other crops suggest a longer size of LD blocks in narrow-based germplasm groups than broad based-germplasm groups in plants [32,33,59], largely influenced by a recombination rate, mating system (selfing vs. out-crossing), genetic isolation, population subdivision and admixture, selection, mutation, and effective population sizes [29]. The genome-wide averages of LD block size for cotton was also longer in the cultivated cotton germplasm group of Mexican and African origin than that of broad-based landrace stock germplasm group, emphasizing the effect of selection pressure, inbreeding and small effective population size in breading history of Upland variety accessions. However, estimates of pairwise LD in variety group might be statistically biased due to small population size.

The genome-wide averages of LD block size for cotton at $r^2 \geq 0.1$ threshold (<10 cM in landrace germplasm and >30 cM improved variety germplasm) was comparable with LD decay estimates reported in some local *Arabidopsis* populations (50–100 cM) [28,30], sugar cane (10 cM) [28] sorghum (50 cM) [60], barley (10–50 cM) [25,62], durum wheat (10–20 cM) [63], and grape (5–10 cM) [64], yet was larger than those reported in *Arabidopsis* (25–250 kb) [28,30], maize (200–400 bp) [58,65] sugar beet (3 cM) [66] and wheat (1–5cM) [67]. However, our estimate of genome-wide averages for the extent of LD

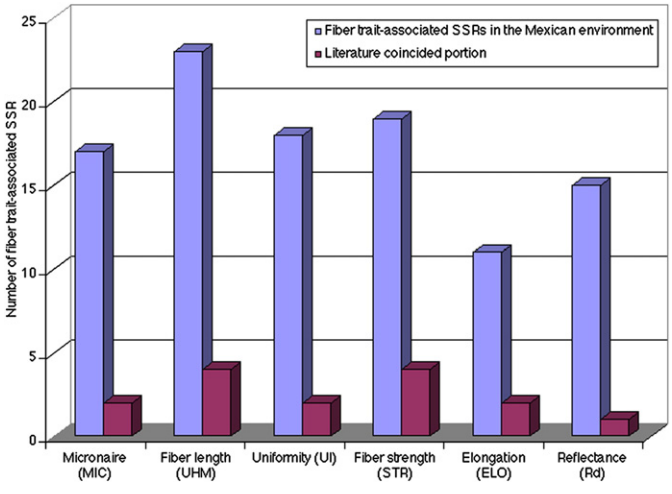

Fig. 4. The results of association mapping in the Mexican environment.

in cotton may not adequately reflect LD patterns of specific regions or specific population groups. It requires additional quantification of LD each targeted region or population group for effective association mapping of variants within regions or populations of interest.

The clear decay of LD with genetic distance reflects the portion of pairwise LD between SSR loci conserved with linkage that is useful for a genome-wide association mapping [27,61] in cotton. However, we observed a number of unlinked markers showing significant LDs between pairs of SSR loci, revealing the existence of other LD generating factors than linkage in the cotton genome [27,28,61]. One of those factors is selection since we observed a greater proportion of unlinked marker pairs in significant LD. This might also be the result of co-selection of loci during domestication process. Relatedness is another factor that might also generate LD between unlinked loci pairs when predominant parents exist in germplasm groups [27,61]. In assayed germplasm about 22% accession pairs had ≥0.1 kinship values, suggesting the potential role of relatedness in observed LD. Theoretically population stratification is one of the main forces generating LD [27,61]. LD generated by selection, population stratification, and genetic drift might be theoretically useful for association mapping (less number of markers are needed) in a specific population groups and situations, yet it tends to reveal spurious marker-trait associations [27,35,61]. This underlies necessity for serious consideration of population structure and relatedness to perform population-based association mapping in cotton germplasm resources, at least, in our samples.

Association mapping of the main fiber quality traits

Similar to the case reported for other crops [25,64] the moderately large LD blocks (<10 cM in landrace germplasm and >30 cM improved variety germplasm) in cotton suggests the potential of conducting effective LD mapping of complex traits with fewer numbers of markers than that of required for the human genome, where a very high density of molecular markers is needed for association mapping in the majority of cases [68,69]. Considering the tetraploid cotton genome with a total recombinational length of about 5200 cM and an average 400 kb per cM [70], the LD block sizes of ~10–30 cM distance in cotton is large enough to conduct an association mapping of complex traits that would require ~200–500 polymorphic markers distributed uniformly across the genome for minimum coverage of LD blocks in the genome of variety germplasm. Further, because of lesser extent of genome-wide LD, perhaps, more markers are needed to effectively scan the genome in landrace stock germplasm for high-resolution mapping. These suggest an opportunity to perform coarse mapping with less number of markers in variety germplasm and fine mapping in landraces stock germplasm, assuming that genetic causations is sufficiently similar across the germplasm groups [71]. This conclusion is preliminary due to the polyploid nature of the Upland cotton genome, and our inability to identify the linkage phase of marker data in our study.

Structured population and relatedness characteristics in assayed cotton germplasm resources suggested a consideration of population structure and kinship in conducting population-based association mapping in cotton germplasm, in particular with our material where predefined groups represented an unbalanced number of accessions [35–37,72]. Hence, we applied the mixed linear model (MLM) approach of Yu et al. [36] considering both population structure (*Q*) and kinship (*K*) to eliminate possible spurious associations, and successfully identified a number of SSR markers significantly associated with the main fiber quality traits of cotton. Compared to our study, several initial association mapping studies in plants used a significantly less number of plant accessions [33,49] to find useful associations. One of the initial association studies in cotton [49] reported SSR marker associations in just 56 accessions of diploid cottons. LD-based association mapping results in 286 exotic cotton accessions should provide useful information for breeding programs of cotton.

Between 6% and 21% of SSRs associated with the main fiber quality traits in our study coincided with reported fiber quality trait-associated SSRs from QTL-mapping studies in various experimental populations, which is useful in the judging of the association results [25]. These markers are useful for MAS program because their validity has been proved across a diverse set of germplasm, broadly covering many meiotic events, and different QTL mapping reports. At the same time, majority of SSR markers associated with fiber quality traits in our diverse cotton germplasm were new markers. This could be due to 1) the use of many new SSR markers from the 'GH' collection (Pepper et al. unpublished) that have not yet been used for mapping of QTLs in experimental populations, or used yet not reported, 2) mapping of potentially new loci contributing to the fiber quality that are absent in current breeding populations. These new markers should also be useful for future MAS programs to mobilize potentially new QTLs from diverse 'underutilized' germplasm resources to elite cultivars.

In conclusion, the results of our study, for the first time, demonstrated the significant potential of LD-based association mapping of complex traits in cotton with a relatively small number of markers. Conversely, the mapping resolution may be limited, in particular, with breeding germplasm. However, a less extent of LD blocks in the exotic *G. hirsutum* germplasm suggest an opportunity to develop a set of mapping populations with the required amount of LD and diversity for high-resolution mapping through directed crossing between the selected germplasm groups [29]. Our results also suggest that population structure and relatedness should be taken into serious account to perform unbiased population-based association mapping in cotton germplasm resources and highlight the potential and feasibility of mixed linear model association mapping in cotton. A number of SSR markers associated with fiber quality traits in diverse exotic cotton germplasm, herein, should be useful to effectively exploit potentially new genetic variations of 'underutilized' wild and primitive *G. hirsutum* races in cotton improvement using MAS programs.

Materials and methods

Plant material

The Uzbek cotton germplasm collection constitutes more than 17,000 cotton germplasm accessions of the A-to K-genome groups from 43 cotton species that have been developed in the Cotton Research Institutes of the Republic of Uzbekistan and collected over the world for the past century [50]. A total of 285 'exotic' accessions, including 208 exotic landrace stocks ('yucatanense', 'punctatum', 'morilli', 'richmondi', 'marie-galante', 'latifolium', 'palmeri' and other unclassified landrace stocks of *G. hirsutum* marked as "wild"), 24 Mexican variety accessions, and 53 African variety accessions from Uzbek cotton germplasm collection were selected for analyses. We also included *G. hirsutum* genetic standard, TM1, and *G. barbadense* standard, 3–79 into a panel. The 77 variety accessions of Mexican and African origin (based on germplasm passport data, unpublished) were included due to their short-day photoperiodic flowering, similar to the landrace stocks, observed in the summer Uzbekistan environment (Table S1).

Phenotypic analyses

Because of their short-day photoperiodic flowering, these *G. hirsutum* accessions were grown in short-day conditions of the Cotton Winter Nursery (CWN) of USDA-ARS, in Tecoman, Mexico in 2005. Cotton accessions are grown in the CWN during the winter dry season under irrigation. Planting consisted of individual plants spaced 18 in. apart in rows spaced 40 in. Seven plants of each accession were grown,

and a pooled fiber sample was analyzed for fiber quality traits (UHM, STR, MIC, ELO, UI, and Rd). Fiber analysis was conducted by the Cotton Incorporated HVI system. The statistical analyses and correlations between traits were performed using the Visual Statistics System (ViSta) [73].

Genotyping with SSR markers

From each accession group, 8–10 young fully expanded leaves were collected and stored at −80 °C. Genomic DNAs were isolated from the frozen leaf tissues using method of Dellaporta et al. [74] with minor modification and optimization for frozen tissues. Prepared genomic DNAs were checked in 0.9% agarose electrophoresis and DNA concentrations were estimated based on Hind III digested λ-phage DNA.

Accessions were genotyped using a 95 core set SSR marker primers (an average ~4 SSR markers per each chromosome covering, 26 chromosomes of cotton; Table S2). This core set of SSRs was chosen specifically for germplasm collection characterization (R. J. Kohel and J. Z. Yu, personal communication) and covered about 18% of the genome, spanning a total of ~950 cM distance or an average of ~43 cM per each chromosome (J. Z. Yu et al. unpublished). PCR-amplifications were performed in a 8 μl reaction mix containing 0.8 μl 10× PCR buffer, 0.2 μl dNTPs (10 mM each), 0.72 μl 25 mM $MgCl_2$, 0.2 μl 5 pM non-labeled, 0.07 μl AmpliTaq Gold DNA polymerase (Applied Biosystems, USA), and 15 ng genomic DNA. PCR amplification was carried out using a PTC-225 DNA Engine Tetrad thermocycler (MJ Research, USA) with first denaturation at 95 °C for 10 min, followed by 10 cycles of 94 °C for 1 min, 60 °C for 1 min (decreases of 0.5 °C in each cycle) and 72 °C for 2 min; 33 cycles of 94 °C for 15 s, 55 °C for 30 s and 72 °C for 1 min. A final 6 min extension at 72 °C was performed. The amplification products were visualized in 3.5% high-resolution agarose with ethiduim bromide and photodocumented using an Alpha Imager (Innotech Inc., USA) gel documentation system.

Molecular genetic diversity and phylogenetic analyses

Considering *G. hirsutum* is an allopolyploid with reticulated germplasm resources and the germplasm material used in this study were strictly self-pollinated during the past 50 years for germplasm renewing, we scored our SSR data like a dominant marker type with "1" for absent, "2" for present state, or "0" (or '?", "−999", and "−9", depending on the software requirement) for the occasional non-amplification or missing data state, taking each band as an independent marker locus with a clear size band separation [67,75] to avoid assigning incorrect allelic relationships.

The heterozygosity level of marker data was identified according to an average similarity frequency of alternative alleles (0.5 vs. 0.5 for high heterozygosity or 0.9 vs. 0.1 low heterozygosity levels) [57] after filtering for minor allele frequency (MAF). Allele frequencies for dominant markers were calculated using SpaGeDi software [76]. The polymorphic information content (PIC) was analyzed using the PowerMarker software package [77]. Genetic distance and phylogenetic analyses of cotton accessions were performed using Neighbor Joining (N-J) algorithms with the minimum evolution objective function [78] of the software package PAUP*4.0b10 [79]. Genetic variation within and among predefined groups and pairwise F_{ST} genetic distances was measured by analysis of molecular variance (AMOVA) [80–82] using ARLEQUIN 2.0 [83]. We also applied a Bayesian method to further study genetic differentiation among population groups, which allows direct estimates of F_{ST} from dominant markers without prior knowledge of inbreeding history [84,85]. Several runs for *full*, $f=0$, $\theta=0$ and $f=free$ models were performed using HICKORY, ver. 1.0, with the default sampling parameters (burn-in = 50,000, sample = 250,000, and thin = 50) following software guidelines [85]. Although the Bayesian analysis with dominant markers revealed the estimates of inbreeding coefficients (F_{IS}) (data not shown), we did not consider or discuss the results of F_{IS} due to the biased nature of the values obtained from small within population sample sizes [85]. In both AMOVA and Bayesian population differentiation analyses, the 5% minor allele filtered SSR dataset was used.

Pairwise linkage disequilibrium and LD decay

For population structure, kinship, pairwise LD and association mapping analyses, only *G. hirsutum* genotypes were used, excluding the control *Gossypium* genotypes, *G. barbadense*. The genome-wide LD between pairs of SSR marker loci was studied according to Witt and Buckler [86] using the software package TASSEL ver. 1.9.6. The genome-wide LD between all pairs of SSR alleles was analyzed with MAF filtered datasets where SSRs alleles with a 0.05 frequency in genotyped accessions were removed before conducting LD analyses because minor alleles are usually problematic and biased for LD estimates between pairs of loci [87,88]. The MAF removal was performed using the TASSEL site filtration function. LD was estimated by a weighted average of squared allele-frequency correlations (r^2) between SSR loci. The significance of pairwise LD (p-values ≤ 0.005) among all possible SSR loci was evaluated using TASSEL with the rapid permutation test in 10,000 shuffles. The LD values between all pairs of SSR loci were plotted as triangle LD plots using TASSEL to estimate the general view of genome-wide LD patterns and evaluate 'block-like' LD structures. Because of small sample sizes within the predefined groups in the cotton germplasm of Mexico and Africa, we did not evaluate a separate pairwise LD for each group to avoid biased estimates. Because of dense coverage, the linkage map information for the 95 core set SSRs was obtained from a new linkage map of RILs derived from the interspecific cross of TM-1 (*G. hirsutum* (AD1)) × 3–79 (*G. barbadense* (AD2)) (J. Z. Yu et al. 2008, unpublished). The r^2 values for pairs of SSR loci were plotted as a function of map distances (cM), and LD decay (at $r^2 < 0.1$) was estimated.

Inference of population structure and kinship

A model-based approach, implemented in the software package STRUCTURE ver. 2.1 [35,72] for dominant markers (coded as 1,−9; 2, −9), was used to identify significant clusters. In the first attempt, we used both 'no-admixture' and 'admixture' co-ancestry models under independent and correlated allele frequencies using the burn-in time of 50,000 and the number of replications at 100,000 [72] with the K up to 10. However, we did not determine distinct clusters and could not assign a significant number of K populations. Therefore, we used the prior population information, pre-defining accessions to a specific type of populations. Accessions were defined as 1) landrace stocks, 2) cultivated variety accessions from Mexico, and 3) cultivated variety accessions from Africa based on the source of origin as indicated in the germplasm collection catalogue (unpublished information). We also analyzed pairwise kinship (K-matrix). Pairwise kinship estimates were calculated according to Hardy and Vekemans [76] using the software package SpaGedi. The kinship coefficient of Hardy [55] was used to obtain the pairwise kinship matrix. The K-matrices (286 by 286), and the Q-matrix describing the assignment of each accession to specific clusters, were used in mixed liner model association mapping.

Association mapping

The LD-based association mapping of fiber quality traits was performed according to Yu et al. [36], using the TASSEL software package. For association mapping, the 5% MAF filtered SSR datasets were used. Fiber trait data was imputed for missing data, and normalized using algorithms implemented in TASSEL before conducting an association mapping.

Data availability

The SSR data genotyped in our germplasm is available to readers through the Cotton Marker Database (CMD; http://www.cottonmarker.org/Downloads.shtml, verified on September 22, 2008; [22]).

Acknowledgments

We acknowledge the Science and Technology Center of Ukraine for the project coordination, and the technical assistance of project participants of P120/P120A. We also thank the administration of Academy of Sciences of Uzbekistan for their continual support of the research efforts. We thank Drs. A. Abdullaev (Uzbek cotton germplasm curator), and S.M. Rizaeva for valuable suggestions on germplasm selection and phenotypic analyses in the Uzbekistan environment. We thank Ms Linda Ballard, Genomics laboratory, USDA-ARS at Mississippi for useful suggestions during manuscript edition.

Funding: This project was funded by the Office of International Research Programs (OIRP) of United States Department of Agriculture (USDA) in the frame of USDA–Former Soviet Union (FSU) cooperation programs with the research grant of P120/P120A to IYA, RJK, JZY, SS, and AEP.

Disclaimer: Mention of trademark or proprietary product does not constitute a guarantee or warranty of the product by the United States Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.07.013.

References

[1] P.A. Fryxell, A revised taxonomic interpretation of *Gossypium* L. (Malvaceae), Rheedea 2 (1992) 108–165.
[2] J.F. Wendel, R.C. Cronn, Polyploidy and the evolutionary history of cotton, Adv. Agron. 78 (2003) 140–186.
[3] G. Sunilkumar, L.M. Campbell, L. Puckhaber, R.D. Stipanovic, K.S. Rathore, Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 18054–18059.
[4] H. Kebede, G. Burow, R.G. Dani, R.D. Allen, A-genome cotton as a source of genetic variability for Upland cotton (*Gossypium hirsutum*), Genet. Resour. Crop Evol. 54 (2007) 885–895.
[5] J.E. Endrizzi, E.L. Turcotte, R.J. Kohel, Genetics, cytology, and evolution of *Gossypium*, Adv. Genet. 23 (1985) 272–375.
[6] L.L. Phillips, Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World × Asiatic and New world × wild American hexaploid, Am. J. Bot. 51 (1964) 324–329.
[7] J.F. Wendel, V.A. Albert, Phylogenetics of the cotton genus (*Gossypium*)—character state weighted Parsimony analysis of chloroplast–DNA restriction site data and its systematic and biogeographic implications, Sys. Bot. 17 (1992) 115–143.
[8] K.L. Adams, R. Percifield, J.F. Wendel, Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid, Genetics 168 (2004) 2217–2226.
[9] J.F. Wendel, C.L. Brubaker, A.E. Percival, Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton, Am. J. Bot. 79 (1992) 1291–1310.
[10] J. Iqbal, O.U.K. Reddy, K.M. El-Zik, A.E. Pepper, A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting, Theor. Appl. Genet. 103 (2001) 547–554.
[11] J.F. Wendel, Cotton, in: S. Simmonds, J. Smart (Eds.), Evolution of Crop Plants, 1st edn., Longman, London, 1995.
[12] C.L. Brubaker, A.H. Paterson, J.F. Wendel, Comparative genetic mapping of allotetraploid cotton and its diploid progenitors, Genome 42 (1999) 184–203.
[13] J.B. Hutchinson, Intraspecific differentiation in *Gossypium hirsutum*, Heredity 5 (1951) 169–193.
[14] J.M. Lacape, D. Dessauw, M. Rajab, J.L. Noyer, B. Hau, Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs, Mol. Breeding 19 (2007) 45–58.
[15] F.M. Mauer, Cotton, Academy of Sciences of USSR Press, Tashkent, 1954 [in Russian].
[16] S. Liu, R.G. Cantrell, J.C.J. McCarty, J.M. Stewart, Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions, Crop Sci. 40 (2000) 1459–1469.
[17] I.Y. Abdurakhmonov, et al., The role of induced mutation in conversion of photoperiod dependence in cotton, J. Hered. 98 (2007) 258–266.
[18] P.L. Hollis, Cotton quality has declined over years. Southeast Farm Press. (2004) http://southeastfarmpress.com/mag/farming_cotton_quality_declined.
[19] G.A. van Esbroeck, D.T. Bowman, O.L. May, D.S. Calhoun, Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars, Crop Sci. 9 (1999) 323–328.
[20] Z.J. Chen, et al., Toward sequencing cotton (*Gossypium*) genomes, Plant Physiol. 145 (2007) 1303–1310.
[21] H. Zhang, et al., Recent advances in cotton genomics, Int. J. Plant Genomics 742304 (2008) doi:10.1155/2008/742304.
[22] A. Blenda, et al., CMD: a cotton microsatellite database resource for *Gossypium* Genomics, BMC Genomics 7 (2006) 132.
[23] K.M. Weiss, A.G. Clark, Linkage disequilibrium and mapping of human traits, Trends Genet. 18 (2002) 19–24.
[24] L. Kruglyak, The road to genome-wide association studies, Nat. Rev. Genet. 9 (2008) 314–318.
[25] A.T.W. Kraakman, R.E. Niks, P.M.M.M. Van den Berg, P. Stam, F.A. Van Eeuwijk, Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars, Genetics 168 (2004) 435–446.
[26] M. Hansen, T. Kraft, S. Ganestam, S. Säll, N.O. Nilsson, Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers, Genet. Res. 77 (2001) 61–66.
[27] B. Stich, et al., Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers, Mol. Breed. 17 (2006) 217–226.
[28] S.A. Flint-Garcia, J.M. Thornsberry, E.S. Buckler IV, Structure of linkage disequilibrium in plants, Annu. Rev. Plant Biol. 54 (2003) 357–374.
[29] A. Rafalski, M. Morgante, Corn and humans: recombination and linkage disequilibrium in two genomes of similar size, Trends Genet. 20 (2004) 103–111.
[30] M. Nordborg, J, et al., The extent of linkage disequilibrium in *Arabidopsis thaliana*, Nat. Genet. 30 (2002) 90–193.
[31] M. Nordborg, et al., The pattern of polymorphism in *Arabidopsis thaliana*, PLoS Biol. 3 (2005) el96.
[32] P.K. Gupta, S. Rustgi, P.L. Kulwal, Linkage disequilibrium and association studies in higher plants: present status and future prospects, Plant Mol. Biol. 57 (2005) 461–485.
[33] I.Y. Abdurakhmonov, A. Abdukarimov, Application of association mapping to understanding the genetic diversity of plant germplasm resources, Int. J. Plant Genomics 574927 (2008) doi:10.1155/2008/574927.
[34] M. Nei, Analysis of gene diversity in subdivided populations, Proc. Natl. Acad. Sci. U. S. A. 70 (1973) 3321–3323.
[35] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[36] J. Yu, et al., A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2) (2006) 203–208.
[37] K. Zhao, et al., An *Arabidopsis* example of association mapping in structured samples, PLoS Genet. 31 (2007) e4.
[38] K. Zhang, M. Deng, T. Chen, M.S. Waterman, F. Sun, A dynamic programming algorithm for haplotype block partitioning, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 7335–7339.
[39] Z.G. Han, W. Guo, X.L. Song, T. Zhang, Genetic mapping of EST-derived microsatellites from the diploid *Gossypium* arboreum in allotetraploid cotton, Mol. Genet. Genomics, 272 (2004) 308–327.
[40] M. Mei, et al., Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*), Theor. Appl. Genet. 108 (2004) 280–291.
[41] J.M. Lacape, et al., QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum*×*Gossypium barbadense* backcross generations, Crop Sci. 45 (2005) 123–140.
[42] X. Shen, et al., Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers, Mol. Breed. 15 (2005) 169–181.
[43] Z.S. Zhang, et al., Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.), Euphytica 144 (2005) 91–99.
[44] D.H. He, et al., Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton, Euphytica 144 (2005) 141–149.
[45] Z. Lin, et al., Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD, Plant Breed. 124 (2005) 180–187.
[46] B. Wang, et al., QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton, Euphytica 152 (2006) 367–378.
[47] I.Y. Abdurakhmonov, et al., Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*, Euphytica 156 (2007) 141–156.
[48] D.H. He, et al., QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* x *Gossypium barbadense*, Euphytica 153 (2007) 181–197.
[49] S.K. Kantartzi, J.M. Stewart, Association analysis of fibre traits in *Gossypium arboreum* accessions, Plant Breed. 127 (2008) 173–179.
[50] I.Y. Abdurakhmonov, Exploiting genetic diversity, in: D. Ethridge (Ed.), Proceedings of World Cotton Research Conference-4: Plenary Presentations and Papers, September 10–14, Lubbock, TX USA, 2007.
[51] M. Ulloa, Heritability and correlations of agronomic and fiber traits in okra-leaf Upland cotton population, Crop Sci. 46 (2006) 1508–1514.
[52] X.X. Ma, et al., Simple sequence repeat genetic linkage maps of A-genome diploid cotton (*Gossypium arboreum*), J. Integrat. Plant Biol. 50 (2008) 491–502.
[53] P.G. Sand, A lesson not learned: allele misassignment, Behav. Brain Funct. 3 (2007) 65.
[54] P.M. Hollingsworth, R.A. Ennos, Neighbor joining trees, dominant markers and population genetic structure, Heredity 92 (2004) 490–498.

[55] O.J. Hardy, Estimation of pairwise relatedness between individuals and characterization of isolation by distance processes using dominant genetic markers, Mol. Ecol. 12 (2003) 1577–1588.
[56] M. Malosetti, C.G. van der Linden, B. Vosman, F.A. van Eeuwijk, A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato, Genetics 175 (2007) 879–889.
[57] H. Iwata, Y. Uga, Y. Yoshioka, K. Ebana, T. Hayashi, Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms, Theor. Appl. Genet. 114 (2007) 1437–1449.
[58] Y. Li, et al., Estimation of multilocus linkage disequilibria in diploid populations with dominant markers, Genetics 176 (2007) 1811–1821.
[59] D.L. Remington, et al., Structure of linkage disequilibrium and phenotypic associations in the maize genome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 11479–11484.
[60] M.T. Hamblin, et al., Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*, Genetics 167 (2004) 471–483.
[61] B. Stich, A.E. Melchinger, M. Frisch, H.P. Maurer, M. Heckenberger, J.C. Reif, Linkage disequilibrium in European elite maize germplasm investigated with SSRs, Theor. Appl. Genet. 111 (2005) 723–730.
[62] L.V. Malysheva-Otto, M.W. Ganal, M.S. Roder, Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.), BMC Genet. 7 (2006) 6.
[63] M. Maccaferri, M.C. Sanguineti, P. Donini, R. Tuberosa, Population structure and long-range linkage disequilibrium in a drum wheat elite collection, Mol. Breed. 15 (2005) 271–289.
[64] A.T. Barnaud, Lacombe, A. Doligez, Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. Theor. Appl. Genet. 112 (2006) 708–716.
[65] M.I. Tenaillon, et al., Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.), Proc. Nat. Acad. Sci. U. S. A. 98 (2001) 9161–9166.
[66] T.M. Kraft, Hansen, N.O. Nilsson, Linkage disequilibrium and fingerprinting in sugar beet, Theor. Appl. Genet. 101 (2000) 323–326.
[67] F. Breseghello, M.E. Sorrells, Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars, Genetics 172 (2006) 1165–1177.
[68] L. Kruglyak, Prospects for whole-genome linkage disequilibrium mapping of common disease genes, Nat. Genet. 22 (1999) 139–144.
[69] D.E. Reich, et al., Linkage disequilibrium in the human genome, Nature 411 (2001) 199–204.
[70] A.H. Paterson, R.H. Smith, Future horizons: biotechnology of cotton improvement, in: C.W. Smith, J.T. Cothren (Eds.), Cotton: Origin, History, Technology, and Production, John Wiley &Sons, Inc., New York, 1999.
[71] D.B. Goldstein, M.E. Weale, Population genomics: linkage disequilibrium holds the key, Curr. Biol. 11 (2001) R576–R579.
[72] J.K. Pritchard, W. Wen, Documentation for Structure Software, The University of Chicago Press, Chicago, 2004.
[73] F.W. Young, C.M. Bann, Data analyses with ViSta, in: J. Fox, R. Stine (Eds.), Statistical Computing Environments for Social Research, Stage, California, 1996, pp. 207–235.
[74] S.L. Dellaporta, J. Wood, J.P. Hicks, A plant DNA mini-preparation: Version II, Plant. Mol. Biol. Rep. 1 (1983) 19–21.
[75] L. Tommasini, T. Schnurbusch, D. Fossati, F. Mascher, B. Keller, Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties, Theor. Appl. Genet. 115 (2007) 697–708.
[76] O.J. Hardy, X. Vekemans, SpaGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels, Mol. Ecol. Notes 2 (2002) 618–620.
[77] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129.
[78] M. Saitou, N. Nei. The neighbor joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[79] D.L. Swofford, Phylogenetic Analysis Using Parsimony (*and other Methods), Sinauer, Sunderland, MA, USA, 2002.
[80] J. Reynolds, B.S. Weir, C.C. Cockerham, Estimation for the co-ancestry coefficient: basis for a short-term genetic distance, Genetics 105 (1983) 7767–7779.
[81] B.S. Weir, C.C. Cockerham, Estimating *F* statistics for the analysis of population structure, Evolution 38 (1984) 1358–1370.
[82] L. Excoffier, P. Smouse, J. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics, 131 (1992) 479–491.
[83] S. Schneider, D. Roessll, L. Excoffier, ARLEQUIN: A Software for Population Genetics Data Analysis, Version 2. 0. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, 2000.
[84] K.E. Holsinger, P.O. Lewis, D.K. Dey, A Bayesian approach to inferring population structure from dominant markers, Mol. Ecol. 11 (2002) 1157–1164.
[85] K.E. Holsinger, P.O. Lewis, HICKORY: A Package for Analysis of Population Genetic Data, Version 1. 0. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA, 2003.
[86] S.R. Witt, E.S. Buckler, Using Natural allelic diversity to evaluate gene function, Methods Mol. Biol. 236 (2003) 123–139.
[87] K.L. Mohlke, et al., Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns, Genome Res. 11 (2001) 1221–1226.
[88] A.F. McRae, et al., Linkage disequilibrium in domestic sheep, Genetics 160 (2002) 1113–1122.